

OSP-8939-42

us

JC530 U.S. PTO

09/510349



日 本 国 特 許 庁  
PATENT OFFICE  
JAPANESE GOVERNMENT

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日

Date of Application:

1999年 2月25日

出 願 番 号

Application Number:

平成11年特許願第049215号

出 願 人

Applicant (s):

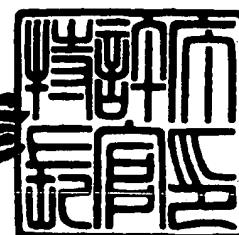
日本電信電話株式会社

CERTIFIED COPY OF  
PRIORITY DOCUMENT

2000年 1月28日

特許庁長官  
Commissioner,  
Patent Office

近 藤 隆 彦



【書類名】 特許願

【整理番号】 NTTH107182

【提出日】 平成11年 2月25日

【あて先】 特許庁長官 殿

【国際特許分類】 H04L 12/00

【発明の名称】 トラヒック観測装置、データグラム転送システムおよび  
データグラム転送方法

【請求項の数】 20

【発明者】

【住所又は居所】 東京都新宿区西新宿三丁目 1 9 番 2 号 日本電信電話株  
式会社内

【氏名】 清水 敬司

【発明者】

【住所又は居所】 東京都新宿区西新宿三丁目 1 9 番 2 号 日本電信電話株  
式会社内

【氏名】 栗本 崇

【特許出願人】

【識別番号】 000004226

【氏名又は名称】 日本電信電話株式会社

【代理人】

【識別番号】 100064908

【弁理士】

【氏名又は名称】 志賀 正武

【手数料の表示】

【予納台帳番号】 008707

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9701417

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 トラヒック観測装置、データグラム転送システムおよびデータグラム転送方法

【特許請求の範囲】

【請求項 1】 現在よりあらかじめ決められた期間前までの、ユーザが網内に送信または網から受信したデータグラムに関するトラヒック情報を観測するトラヒック観測機能と、

該トラヒック観測機能により得られたトラヒック情報からデータグラムの送出に対する網の評価により得られるプレファレンス値を計算するプレファレンス値計算機能と、

該プレファレンス値計算機能により計算されたプレファレンス値を評価値として前記送信されるデータグラムのヘッダに書き込むプレファレンス値挿入機能とを備えたことを特徴とするトラヒック観測装置。

【請求項 2】 ユーザ端末から送信されたデータグラムを受けたデータグラム転送装置が、そのデータグラムのヘッダに記載された転送宛先アドレスに向けて前記データグラムをリレーするデータグラム転送システムにおいて、前記データグラムに関するトラヒック情報に基づいて網に対するインパクトを評価し、その評価値を前記ヘッダに書き込んで、プレファレンス値をつけるトラヒック観測手段を設けたことを特徴とするデータグラム転送システム。

【請求項 3】 前記データグラムのヘッダに記載された前記評価値に従った順序によりデータグラムの転送処理を行う手段を有することを特徴とする請求項 2 に記載のデータグラム転送システム。

【請求項 4】 前記データグラム転送装置が、前記データグラムをいずれかの入力インタフェース部からいずれかの出力インタフェース部へ競合することなく転送するバックプレーンスイッチ部を有し、

前記出力インタフェース部が、受信したデータグラムのヘッダから前記プレファレンス値を取得し、そのプレファレンス値の小さいものから順に、輻輳状態としない範囲で転送すべきデータグラムを選択して、優先的にバッファメモリに書き込むバッファ書き込み制御部を有することを特徴とする請求項 2 に記載のデ

ータグラム転送システム。

【請求項 5】 前記トラヒック情報が、データグラムのサイズまたは連続して送出されるデータグラムの時間間隔であることを特徴とする請求項 2 に記載のデータグラム転送システム。

【請求項 6】 前記トラヒック観測手段が、前記プレファレンス値としてデータグラムのヘッダに存在するデータグラムのサイズフィールドを利用することを特徴とする請求項 2 に記載のデータグラム転送システム。

【請求項 7】 前記トラヒック観測手段が、一つ前の前記データグラムの送出時刻と現在時刻との差の逆数およびこれらの各時刻に送出および到着したデータグラムの大きさに基づいて得られる値をプレファレンス値として計算することを特徴とする請求項 2 に記載のデータグラム転送システム。

【請求項 8】 前記トラヒック観測手段が、前記データグラムのサイズおよび連続するデータグラムの間隔からスライディングウィンドウ方式による平均レートをプレファレンス値として計算することを特徴とする請求項 2 に記載のデータグラム転送システム。

【請求項 9】 前記トラヒック観測手段が、前記データグラムのサイズおよび連続するデータグラムの間隔から観測期間の平均レートをプレファレンス値として計算することを特徴とする請求項 2 に記載のデータグラム転送システム。

【請求項 10】 前記トラヒック観測手段が、ユーザが送出したデータグラムの個数と受信したデータグラムの個数との差をプレファレンス値として求めることを特徴とする請求項 2 に記載のデータグラム転送システム。

【請求項 11】 前記バッファ書き込み制御部が、受信したデータグラムのプレファレンス値をプレファレンス値取り出し機能により取得し、このプレファレンス値をキーにしてプレファレンス比較機能にてソーティングを行い、データグラムにプレファレンス値の昇順に順序付けを行い、さらにその順序に従って書き込み制御機能にバッファメモリに対する書き込み処理を行わせることを特徴とする請求項 4 に記載のデータグラム転送システム。

【請求項 12】 前記バッファメモリが複数の優先順位を持ったクラス別バッファメモリ部を有し、前記バッファ書き込み制御部がプレファレンス値に従っ

てその優先順位を持った複数のクラスを選択し、前記データグラムの書き込みを行うことを特徴とする請求項4に記載のデータグラム転送システム。

【請求項13】 前記プレファレンス値の取得処理が、データグラムの到着率に応じて変更可能な可変間隔または常に固定の固定間隔の期間ごとに行われることを特徴とする請求項11に記載のデータグラム転送システム。

【請求項14】 前記バッファ書き込み制御部では、データグラムの前記バッファメモリへの書き込み前に転送判断を行い、転送を行わないと判断した場合には、前記バッファメモリに空き領域があっても廃棄を行い、転送を行うと判断した場合には前記バッファメモリへの書き込み処理を行うことを特徴とする請求項11に記載のデータグラム転送システム。

【請求項15】 前記バッファ書き込み制御部が、既にバッファメモリに書き込んである処理済みデータグラムのプレファレンス値の総和を計算し、この計算結果に基づく確率計算を行って、その確率に応じてデータグラムの廃棄を行うことを特徴とする請求項11に記載のデータグラム転送システム。

【請求項16】 前記バッファ書き込み制御部が、データグラムにプレファレンス値の昇順に順序付けを行った後、その順序付けに従ったバッファメモリの使用量の概算値に基づく確率を求め、得られた確率に応じてデータグラムの廃棄を行うことを特徴とする請求項11に記載のデータグラム転送システム。

【請求項17】 前記バッファ書き込み制御部が、適当なタイミングにてスレッシュホールド値を計算しておき、到着したデータグラムのプレファレンス値と前記スレッシュホールド値との比較を行い、プレファレンス値がスレッシュホールド値より大きい場合に前記データグラムを廃棄し、小さい場合に前記バッファメモリへの書き込みを行うことを特徴とする請求項11に記載のデータグラム転送システム。

【請求項18】 前記バッファ書き込み制御部が、適当なタイミングにてスレッシュホールド値を計算しておき、到着したデータグラムのプレファレンス値と前記スレッシュホールド値との比較を行い、これらの各値の差を入力とする関数により確率を計算し、得られた確率に応じてデータグラムの廃棄を行うことを特徴とする請求項11に記載のデータグラム転送システム。

【請求項 19】 前記バッファ書き込み制御部が、あらかじめ決められた範囲のデータグラムへの到着、転送および廃棄のうちのいずれかの事象が発生した時刻および前記プリアレンス値を保存するプリアレンス値保存機能を有し、これらのプリアレンス値を用いて前記スレッシュホールド値が計算されることを特徴とする請求項 17 または請求項 18 に記載のデータグラム転送システム。

【請求項 20】 ユーザによるデータグラムの送出時に、そのデータグラムが網に対してどのようなインパクトを与えるかという評価を行い、この評価を反映したプリアレンス値を計算して、その計算結果を前記データグラムのヘッダに書き込み、前記プリアレンス値による順序付けに従って前記データグラムの転送処理を行うことを特徴とするデータグラム転送方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

この発明は、高速コンピュータ間通信を提供する公衆網におけるベストエフォート型データ通信サービスを提供するためのトラフィック観測装置、データグラム転送システムおよびデータグラム転送方法に関する。

【0002】

【従来の技術】

インターネットで用いられているインターネットプロトコル (IP) パケットのように、その配達に網によって保証されないデータユニットをデータグラムと呼ぶ。このようなデータグラムの転送によりデータ通信サービスを実現する網をデータグラム転送網と呼ぶ。このようなデータグラム転送網では、網内のデータグラム転送装置 (例えば、インターネットではルータと呼ばれる装置が相当する) が、任意の大きさを持ったデータグラムをそのヘッダに記載された宛先アドレスに向かってリレーすることにより、データ通信サービスが実現される。また、このデータグラム転送装置では、自らの処理能力を超えるデータグラムがごく短い期間に集中して到着した場合 (これを輻輳状態と呼ぶ)、到着したデータグラムを適宜廃棄してしまう。このため、データグラム転送網では、「網は個々のデータグラムをその宛先への配達を目指し、網内の転送装置の能力の可能な限りの

転送を行う」というベストエフォート型データ通信サービスを提供することとなる。

#### 【0003】

図12は、これまで述べたようなデータグラム転送システムを構成するデータグラム転送装置の一般的な構成を示すブロック図であり、このデータグラム転送装置は、複数の入力インターフェース部（以下、入力I/F部という）1と複数の出力インターフェース部（以下、出力I/F部という）2とが、一つのバックプレーンスイッチ部3に接続された構成をとっている。このバックプレーンスイッチ部3はいずれか任意の入力I/F部1からいずれか任意の出力I/F部2へ、内部で競合することなくデータグラムを転送する能力を持っている。

#### 【0004】

図13は、図12における入力I/F部1を詳細に示す。この入力I/F部1は、ラインI/F部1a、転送宛先テーブル1b、転送処理部1cおよびデータグラム転送部1dから構成されている。そして、この入力I/F部1では、ラインI/F部1aにおいて入力リンクからのデータグラムが受信されると、転送処理部1cは転送宛先テーブル1bを参照し、宛先アドレスから所望の出力I/F部12を決定するとともに、前記バックプレーンスイッチ部3を介して受信したデータグラムを所望の出力I/F部2へ転送させるために、適当なデータグラム転送部1dへデータグラムを送る。

#### 【0005】

図14は、図12における出力I/F部2を詳細に示す。この出力I/F部2は、データグラム受信部2a、バッファメモリ2b、バッファ書き込み制御部2c、バッファ読み出し制御部2dおよびラインI/F部2eから構成されている。そして、この出力I/F部2では、バックプレーンスイッチ部3よりデータグラム受信部2aに受信されたデータグラムは、バッファ書き込み制御部2cの処理方法に従って、バッファメモリ2bに書き込まれ、出力リンクが利用可能になるのを待つ。すなわち、図15のフローチャートに示すように、あらかじめ決められた順（通常はI/F番号の昇順など）に入力I/F部1を選択し（ステップS1）、そこから到着しているデータグラムの有無を調べる（ステップS2）。



到着している場合には、これをバッファメモリ 2 b の空き領域に書き込み（ステップ S 3）、未処理の入力 I / F 部 1 があれば（ステップ S 4）、ステップ S 1 以下の処理を繰り返し、なければ終了する。これらの入力 I / F 部 1 に対する処理は、出力リンクにおける処理単位時間（通常は一つのデータグラムを送出する時間）に対して入力 I / F 部 1 の数が n の場合、n 倍の速度で行う必要がある。一方、バッファ読み出し制御部 2 d は、ライン I / F 2 e が空きになると、書き込まれた順にバッファメモリ 2 b からライン I / F 部 2 e へデータグラムを転送する。そして、このデータグラムはライン I / F 部 2 e により出力リンクへ送信される。

## 【0006】

従来のデータグラム転送装置では、この出力 I / F 部 2 におけるバッファ書き込み制御部 2 d の処理方法として、バッファメモリ 2 b に空きがある限り、到着した順にデータグラムを書き込んで行くという方法が採用されている。この処理方法とバッファ読み出し制御部 2 d にて書き込まれた順に読み出す方法とを組み合わせた方法は、最初に到着したデータグラムが最初に出力されることから、ファーストインファーストアウト（FIFO）方式と呼ばれる。

## 【0007】

また、バッファメモリ 2 b に空きがない期間にデータグラムが到着した場合、通常それらはすべて廃棄されるが、特にインターネットにおいては、バッファメモリ 2 b を使い尽くす状況を回避し、多くのデータグラムが一度に廃棄されるという状況を防ぐことで、リンクの利用効率を高めることが可能となる。そのため、インターネットで用いられるルータでは、バッファ書き込み方法として、バッファメモリにたとえ空きがあったとしても、バッファメモリの使用量に応じた適当な確率であらかじめデータグラムを廃棄してしまうというランダムアーリーデテクション（RED）方式が採用されている。

## 【0008】

また、図 16 はこの RED 方式による書き込み処理手順を示す。これは、あらかじめ決められた順に入力 I / F 部 1 を選択し（ステップ S 1）、そこから到着しているデータグラムの有無を調べ（ステップ S 2）、バッファメモリ 2 b の使

用量の概算値を求め（ステップ S 5）、この概算値に基づいた確率を求め（ステップ S 6）、続いて転送判断を行い（ステップ S 7）、この判断結果に基づき、転送を行わない場合にはバッファメモリ 2 b に空きがあっても廃棄を行うという手順をとる。

#### 【0009】

このように、従来のデータグラム転送装置にあっては、バッファメモリ 2 b への書き込み方法に注目すると、F I F O 方式の場合はデータグラムの到着順に、一方、R E D 方式の場合でも、バッファメモリ 2 b の使用量に応じた適当な確率でパケットを廃棄するものの、基本的には到着順にバッファメモリ 2 b へデータグラムを書き込んでいる。そして、書き込み時にバッファメモリ 2 b に空きがない場合にはデータグラムを廃棄するというものであり、また、データグラムを転送するか廃棄するかの判断を、データグラムそのものに関する情報ではなく、出力 I / F 部 2 での情報、つまりバッファメモリ 2 b の利用状況、例えば、バッファメモリに空きがない、バッファメモリの空きが 10 % を下回っている等に基づいて行っている。

#### 【0010】

##### 【発明が解決しようとする課題】

しかしながら、このような方法を用いる従来のデータグラム転送システムにおいては、データグラム転送装置がデータグラムの転送処理を出力 I / F 部 2 で得られる情報のみしか考慮に入れずデータグラムの到着順に行っており、短い期間に他のユーザより多くのデータグラムを送信することで、送信したデータグラムの廃棄も多くなるが、全体としては他のユーザより多くの網資源を獲得するということが可能となる。多くのユーザがより多くの網資源を獲得することを目的として、このように多くのデータグラムを送出した場合、網内のそれぞれのデータグラム転送装置は輻輳状態に陥り、送出されたデータグラムの多くを廃棄することとなる。この結果、廃棄されずに最終的な宛先まで到着するデータグラムの数が急激に減少し、網全体としての実効的なデータ転送能力が低下した状態、つまり輻輳崩壊状態に陥ってしまうという課題があった。

#### 【0011】

この発明は前記課題を解決するものであり、データグラム転送装置の出力 I/F 部におけるバッファ書き込み制御部が、データグラムの送出に対する網の評価により得られる情報（プレファレンス値）を利用し、データグラムを到着順以外の方法で転送処理を行うことで、ユーザがより多くの網資源を獲得しようとして必要以上のデータグラムを送出することを抑制でき、これにより輻輳崩壊状態に陥らない安定したデータグラム通信網を実現できるデータグラム転送システムを得ることを目的とする。

## 【0012】

## 【課題を解決するための手段】

前記目的達成のため、請求項 1 の発明にかかるトラヒック観測装置は、現在よりあらかじめ決められた期間前までの、ユーザが網内に送信または網から受信したデータグラムに関するトラヒック情報を観測するトラヒック観測機能と、該トラヒック観測機能により得られたトラヒック情報からデータグラムの送出に対する網の評価により得られるプレファレンス値を計算するプレファレンス値計算機能とを有し、プレファレンス値挿入機能に、前記プレファレンス値計算機能により計算されたプレファレンス値を前記評価値として前記送信されるデータグラムのヘッダに書き込ませるようにしたものである。

## 【0013】

また、請求項 2 の発明にかかるデータグラム転送システムは、ユーザ端末から送信されたデータグラムを受けたデータグラム転送装置が、そのデータグラムのヘッダに記載された転送宛先アドレスに向って前記データグラムをリレーするデータグラム転送システムに、前記データグラムに関するトラヒック情報に基づいてユーザのトラヒックの網に対するインパクトを評価し、その評価値を前記ヘッダに書き込んで、この評価に従った順序による転送処理を行わせるトラヒック観測手段を設けたものである。

## 【0014】

また、請求項 3 の発明にかかるデータグラム転送システムは、前記データグラムのヘッダに記載された前記評価値に従った順序によりデータグラムの転送処理を行う手段を設けたものである。

## 【 0 0 1 5 】

また、請求項 4 の発明にかかるデータグラム転送システムは、前記データグラム転送装置に、前記データグラムをいずれかの入力インタフェース部からいずれかの出力インタフェース部へ競合することなく転送するバックプレーンスイッチ部を設け、前記出力インタフェース部が、受信したデータグラムのヘッダから前記プレファレンス値を取得し、そのプレファレンス値の小さいものから順に、輻輳状態とならない範囲で転送すべきデータグラムを選択して、優先的にバッファメモリに書き込むバッファ書き込み制御部を設けたものである。

## 【 0 0 1 6 】

また、請求項 5 の発明にかかるデータグラム転送システムは、前記トラヒック情報を、データグラムのサイズまたは連続して送出されるデータグラムの時間間隔としたものである。

## 【 0 0 1 7 】

また、請求項 6 の発明にかかるデータグラム転送システムは、前記トラヒック観測手段に、前記プレファレンス値としてデータグラムのヘッダに存在するデータグラムのサイズフィールドを利用させるようにしたものである。

## 【 0 0 1 8 】

また、請求項 7 の発明にかかるデータグラム転送システムは、前記トラヒック観測手段に、一つ前の前記データグラムの送出時刻と現在時刻との差の逆数をプレファレンス値として計算させるようにしたものである。

## 【 0 0 1 9 】

また、請求項 8 の発明にかかるデータグラム転送システムは、前記トラヒック観測手段に、前記データグラムのサイズおよび連続するデータグラムの間隔からスライディングウィンドウ方式による平均レートをプレファレンス値として計算させるようにしたものである。

## 【 0 0 2 0 】

また、請求項 9 の発明にかかるデータグラム転送システムは、前記トラヒック観測手段に、前記データグラムのサイズおよび連続するデータグラム間隔から観測期間の平均レートをプレファレンス値として計算させるようにしたものであ

る。

【 0 0 2 1 】

また、請求項 1 0 の発明にかかるデータグラム転送システムは、前記トラヒック観測手段に、ユーザが送出したデータグラムの個数と受信したデータグラムの個数との差をプレファレンス値として求めさせるようにしたものである。

【 0 0 2 2 】

また、請求項 1 1 の発明にかかるデータグラム転送システムは、前記バッファ書き込み制御部で、受信したデータグラムのプレファレンス値をプレファレンス値取り出し機能により取得し、このプレファレンス値をキーにしてプレファレンス比較機能にてソーティングを行い、データグラムにプレファレンス値の昇順に順序付けを行い、さらにその順序に従って書き込み制御機能にバッファメモリに対する書き込み処理を行わせるようにしたものである。

【 0 0 2 3 】

また、請求項 1 2 の発明にかかるデータグラム転送システムは、前記バッファメモリに複数の優先順位を有するクラス別バッファメモリ部を持たせ、前記バッファ書き込み制御部にプレファレンス値に従ってその優先順位を持った複数のクラスを選択し、前記データグラムの書き込みを行わせるようにしたものである。

【 0 0 2 4 】

また、請求項 1 3 の発明にかかるデータグラム転送システムは、前記プレファレンス値の取得処理を、データグラムの到着率に応じて変更可能な可変間隔または常に固定の固定間隔の期間ごとに行うようにしたものである。

【 0 0 2 5 】

また、請求項 1 4 の発明にかかるデータグラム転送システムは、前記バッファ書き込み制御部で、データグラムの前記バッファメモリへの書き込み前に転送判断を行わせ、転送を行わないと判断した場合には、前記バッファメモリに空き領域があっても廃棄を行わせ、転送を行うと判断した場合には前記バッファメモリへの書き込み処理を行わせるようにしたものである。

【 0 0 2 6 】

また、請求項 1 5 の発明にかかるデータグラム転送システムは、前記バッファ

書き込み制御部で、既にバッファメモリに書き込んである処理済みデータグラムのプレファレンス値の総和を計算させ、この計算結果に基づく確率計算を行わせて、その確率に応じてデータグラムの廃棄を行わせるようにしたものである。

## 【 0 0 2 7 】

また、請求項 1 6 の発明にかかるデータグラム転送システムは、前記バッファ書き込み制御部で、データグラムにプレファレンス値の昇順に順序付けを行わせた後、その順序付けに従ったバッファメモリの使用量の概算値に基づく確率を求めさせ、得られた確率に応じてデータグラムの廃棄を行わせるようにしたものである。

## 【 0 0 2 8 】

また、請求項 1 7 の発明にかかるデータグラム転送システムは、前記バッファ書き込み制御部で、適当なタイミングにてスレッシュホールド値を計算しておき、到着したデータグラムのプレファレンス値と前記スレッシュホールド値との比較を行わせ、プレファレンス値がスレッシュホールド値より大きい場合に前記データグラムを廃棄させ、小さい場合に前記バッファメモリへの書き込みを行わせるようにしたものである。

## 【 0 0 2 9 】

また、請求項 1 8 の発明にかかるデータグラム転送システムは、前記バッファ書き込み制御部にプレファレンス値とスレッシュホールド値の各値の差を入力とする関数により確率を計算させ、得られた確率に応じてデータグラムの廃棄を行わせるようにしたものである。

## 【 0 0 3 0 】

また、請求項 1 9 の発明にかかるデータグラム転送システムは、前記バッファ書き込み制御部に、あらかじめ決められた範囲のデータグラムの到着、転送および廃棄のうちのいずれかの事象が発生した時刻および前記プレファレンス値を保存するプレファレンス値保存機能を持たせ、これらのプレファレンス値を用いて前記スレッシュホールド値を計算させるようにしたものである。

## 【 0 0 3 1 】

また、請求項 2 0 の発明にかかるデータグラムの転送方法は、ユーザによるデ

ータグラムを送出時に、そのデータグラムが網に対してどのようなインパクトを与えるかという評価を行い、この評価を反映したプレファレンス値を計算して、その計算結果を前記データグラムのヘッダに書き込み、前記プレファレンス値による順序付けに従って前記データグラムの転送処理を行うようにしたものである。

### 【0032】

#### 【発明の実施の形態】

以下、この発明の実施の一形態を図について説明する。この発明のデータグラム転送システムで実行されるプレファレンス値を用いた選択的データグラム転送方法は、以下の2点を組み合わせて実行される。

1. 網はユーザのデータグラム送出時にそのデータグラムが網に対しどのようなインパクト（負荷）を与えるかという評価を行い、それを反映させた値としてプレファレンス値を計算し、それをデータグラムのヘッダに付与する。ここで、網に対してのインパクトが少ないものに対して、小さいプレファレンス値を与える。

2. 網内のデータグラム転送装置においては、データグラムの到着順ではなく、このプレファレンス値による順序付けに従った順にデータグラムの転送処理を行う。具体的には網に対してのインパクトが少ないことを意味するプレファレンス値の小さいものほど高い確率で優先して転送を行うよう制御する。

### 【0033】

図13は前記データグラムの転送方法を実行するデータグラム転送システムを示すブロック図であり、ここでは一般的な網とユーザ端末との接続形態の様子とこの発明を実現する装置の一つであるトラヒック観測装置が導入される位置を具体的に示してある。これによればユーザ端末4からライン終端装置5およびアクセス網6を介して送信されたデータグラムは、網内のデータグラム転送装置Dにおいて、図1に示すような入力I/F部1の転送宛先テーブル1bに従って、宛先アドレスの示す端末までリレーされて行く。このアクセス網を終端するライン終端装置7と網内の最初のデータグラム転送装置Dとの間にトラヒック観測手段としてのトラヒック観測装置8が接続されている。

## 【 0 0 3 4 】

図 2 はこのトラヒック観測装置 8 の構成を示し、これがトラヒック観測機能 8 a、プレファレンス値計算機能 8 b およびプレファレンス値挿入機能 8 c により構成され、これらのうちトラヒック観測機能 8 a は、現在よりあらかじめ決められた期間前までの、ユーザ網内に送信、あるいは網より受信したデータグラムに関するトラヒック情報、例えばデータグラムのサイズや連続して送出されたデータグラムの時間間隔などを観測し、保持する機能である。また、プレファレンス値計算機能 8 b は、トラヒック観測機能 8 a により保持されている情報から決められた計算式を用いて、プレファレンス値を計算する機能である。また、プレファレンス値挿入機能 8 c は、到着したデータグラムに、プレファレンス値計算機能 8 b により計算された値をデータグラムのヘッダに書き込む機能である。すなわち、このようなトラヒック観測装置 8 では、トラヒック観測機能 8 a により得られたデータを用いて、プレファレンス値計算機能 8 b により、ユーザのトラヒックの網に対するインパクト（負荷）を評価してその数値化を行い、それをデータグラムのヘッダに書き込むことを実現する。そして、前記トラヒック観測装置 8 のプレファレンス値挿入機能 8 c により書き込まれるプレファレンス値は、図 3 に示すように、転送されるデータグラムのヘッダ部分におけるプレファレンス値を書き込むフィールドに保存する。このフィールドの値は、データグラム転送網内を転送される間変更されず、各データグラム転送装置 D に、網がトラヒック観測装置 8 において行った評価結果を伝える機能を実現する。

## 【 0 0 3 5 】

図 4 は、この発明のデータグラム転送装置 D の出力 I / F 部 2 におけるバッファ書き込み制御部 2 c の構成を示し、これがプレファレンス値取り出し機能 2 c 1、プレファレンス値比較機能 2 c 2 および書き込み制御機能 2 c 3 から構成される。これらのうち、プレファレンス値取り出し機能 2 c 1 は、到着したデータグラムのヘッダよりプレファレンス値を取り出して保持し、プレファレンス値比較機能 2 c 2 は、保持されている各プレファレンス値の比較を行い、到着したデータグラムに対しプレファレンス値のより小さいものが先になるような順序を与える。さらに、書き込み制御機能 2 c 3 は、プレファレンス値比較機能 2 c 2 の



結果を用い、プレファレンス値の小さいものから順に、輻輳状態とならない範囲で転送するデータグラムを選択し、それらを図3に示すようなバッファメモリ2bへ書き込む制御を行い、また、選択されなかったデータグラムは廃棄する制御を行う。

## 【0036】

図5は、この発明における書き込み制御機能2c3による書き込み処理方法を示すフローチャートである。これは、従来のFIFO方式とは違い、最初にプレファレンス値取り出し機能2c1により到着しているデータグラムのプレファレンス値のリストを取得する（ステップS11）。次に、プレファレンス値比較機能2c2により、この値をキーにしソーティングを行い（ステップS12）、データグラムにプレファレンス値の昇順に順序付けを行う（ステップS13）。そして、その順序に従って、書き込み制御機能2c3によりバッファメモリ2bへの書き込み処理を行い（ステップS14）、未処理のデータグラムがある場合には（ステップS15）、ステップS13以下の処理を実行し、ない場合には処理を終了する。

## 【0037】

この発明では、データグラムはデータグラム転送装置Dにおいて、到着順に処理されるのではなく、ヘッダに存在するプレファレンス値の小さいものが優先的に転送されることとなる。このプレファレンス値は、網がユーザのトラヒックを観測することで付与するため、たとえ、あるユーザが短い期間に多くのデータグラムを送信したとしても、それらのデータグラムにはトラヒック観測装置が大きなプレファレンス値をつけるようになり、データグラム転送装置Dでより低いプレファレンスをもつデータグラムよりも高い確率で廃棄されてしまう。従って、短い期間に多くのデータグラムを送信することでデータグラムの廃棄も多くなるが、結果的により多くの網資源を獲得することが不可能となる。その結果、ユーザが多くの網資源の獲得を目的として不必要なデータグラムを送出することを抑制でき、データグラム転送網が輻輳崩壊状態に陥る危険性を低くすることが可能となる。

## 【0038】

前記のプレファレンス値を用いた選択的データグラム転送方法は、網によるユーザトラヒックの評価とその数値化およびその数値による順序付けによる優先転送処理の二つの機能的要素からなり、これらの要素は、それぞれトラヒック観測装置 8 におけるプレファレンス値計算機能 8 b およびデータグラム転送装置 D におけるバッファ書き込み制御部 2 c の処理方法として実施されるものである。以下では、これらの二つについての実施例をそれぞれ述べる。

## 【 0 0 3 9 】

前記のように、バッファ書き込み制御部 2 c では、最初にプレファレンス値取り出し機能 2 c 1 により到着しているデータグラムのプレファレンス値のリストを取得し、プレファレンス比較機能として、このプレファレンス値をキーにしてソーティングを行い、データグラムにプレファレンス値の昇順に順序付けを行い、さらにその順序に従って、バッファメモリ 2 b への書き込み処理を行っている。そしてプレファレンス値の取得処理は、適当な期間 T 毎に行う。この期間 T を長くすれば、その間に到着してソーティングの対象となるデータグラムの数が増加するため、多くのデータグラム間で優先順位の比較ができ、この発明の効果は大きくなるが、到着してから転送処理が行われ、出力リンクへ送信されるまでの遅延時間が大きくなる。この期間 T は、その点を考慮して最適な値を選択する。また、常に固定した期間でもよいが、データグラムの到着率に応じて適応的に期間を変更することも可能である。

## 【 0 0 4 0 】

また、図 6 に示すように、図 5 の処理手順に加えてデータグラムのバッファメモリ 2 b への書き込みをする前に、転送判断を別途行い（ステップ S 1 6）、この判断結果に基づき、転送を行わない場合には（ステップ S 1 7）、バッファメモリ 2 b に空き領域があっても廃棄を行うという処理手順をとることも可能である。一つのより具体的な例を図 7 に示す。既に処理を行いバッファメモリに書き込みを行ったデータグラムのプレファレンス値の和を計算し（ステップ S 1 8）、それに基づいた確率を求め（ステップ S 1 9）、その確率に応じて廃棄を行うことが可能となる。この場合、後に処理されるデータグラム、つまりプレファレンス値の大きいデータグラムほど高い確率で廃棄されることになり、この処理を

行わない図 5 の場合に比べてこの発明の効果は大きくなる。また、この発明の書き込み処理方法は、インターネットのルータに用いられている前記 R E D 方式との併用も可能である。図 8 はこの R E D 方式を併用した書き込み処理手順を示すフローチャートである。これによれば、最初にプレファレンス値のリストを取得し（ステップ S 4 0）、ソーティングを行うことで（ステップ S 1 2）、得られるデータグラムの順序付けに従って（ステップ S 1 3）、R E D 方式の転送処理手順が行われる。すなわち、バッファメモリ 2 b の使用量の概算を行い（ステップ S 2 0）、この概算値に基づいた確率を求め（ステップ S 2 1）、続いて転送判断を行った後（ステップ S 1 7）、その判断結果に従ってバッファメモリ 2 b への書き込みまたは廃棄を行うという手順をとる。

#### 【 0 0 4 1 】

ところで、図 6 ～ 図 8 の例では、プレファレンス値の取得処理を行う期間 T を適当に設定する必要がある。それに対して、図 9 に示す例では、到着したデータグラムは前記のように同時に到着したものが一度に処理されるのではなく、順番に一つずつ処理され、プレファレンス値を取得した後ソーティングを行うという処理を行わない。すなわち、あらかじめ決められた順に入力 I / F 部 1 を選択し（ステップ S 3 1）、そこから到着しているデータグラムの有無を調べ（ステップ S 3 2）、続いて図 1 0 に示すようなプレファレンス値取り出し機能 2 c 1 にてプレファレンス値の取得を行って（ステップ S 3 3）、このプレファレンス値の保存をプレファレンス値保存機能 2 c 4 に行う。さらにスレッシュールド計算機能 2 c 5 において、適当なタイミングにてスレッシュールド値を計算しておき（ステップ S 3 4）、到着したデータグラムのプレファレンス値とスレッシュールド値との比較を行い（ステップ S 3 5）、プレファレンス値がスレッシュールド値より大きい場合は、廃棄され、そうでなければ転送を行うという方法で、選択的なバッファ書き込み処理を行う（ステップ S 1 4）。この例ではデータグラムの選択順序を変える必要がないため、従来法からの改良が比較的簡単であるという特徴を持つ。なお、プレファレンス値とスレッシュールド値の比較を行う場合に、プレファレンス値とスレッシュールド値の差を入力とする関数により確率を計算し（ステップ S 3 5）、その確率によりデータグラムを廃棄する（ステップ S 1 7）と

いう方法も採用可能である。

#### 【 0 0 4 2 】

この発明におけるデータグラム転送装置 D におけるバッファ書き込み制御部 2 c は、あらかじめ決められた範囲のデータグラムの到着、転送、廃棄等のいずれかの事象が起った時刻と、そのプレファレンス値を保存する機能を具備している。また、前記スレッシュホールド値は、それらの各値を用いて以下に示すような値として計算する。

1. 転送した過去 n 個のパケットに関する平均値／メジアン／平均＋分散／平均＋標準偏差／移動平均／分布の最頻値
2. 到着した過去 n 個のパケットに関する平均値／メジアン／平均＋分散／平均＋標準偏差／移動平均／分布の最頻値
3. 廃棄された過去 n 個のパケットに関する平均値／メジアン／平均＋分散／平均＋標準偏差／移動平均／分布の最頻値
4. 過去 t 秒以内に転送したパケットに関する平均値／メジアン／平均＋分散／平均＋標準偏差／移動平均／分布の最頻値
5. 過去 t 秒以内に到着したパケットに関する平均値／メジアン／平均＋分散／平均＋標準偏差／移動平均／分布の最頻値
6. 過去 t 秒以内に廃棄したパケットに関する平均値／メジアン／平均＋分散／平均＋標準偏差／移動平均／分布の最頻値
7. 過去 t 秒以内に転送した、過去 n 個のパケットに関する平均値／メジアン／平均＋分散／平均＋標準偏差／移動平均／分布の最頻値
8. 過去 t 秒以内に到着した、過去 n 個のパケットに関する平均値／メジアン／平均＋分散／平均＋標準偏差／移動平均／分布の最頻値
9. 過去 t 秒以内に廃棄した、過去 n 個のパケットに関する平均値／メジアン／平均＋分散／平均＋標準偏差／移動平均／分布の最頻値

#### 【 0 0 4 3 】

次に、出力 I / F 部 2 に複数のクラス別バッファメモリ部を持つバッファメモリ 2 b を図 1 1 に示す。これらのクラス別バッファメモリ部はそれぞれ優先順位をもっており、その優先順位に従ってバッファ読み出し制御部 2 d により読み出

しが行われる。バッファ書き込み制御部 2 c は、プレファレンス値比較機能により与えられた順位に従って、複数の優先順位を持ったクラス別バッファメモリ部の選択を行い、そこにデータグラムを書き込む。具体的には以下の処理をバッファ書き込み制御部 2 c において行う。すなわち、いま、 $n$  個の優先順位を持ったキューが存在し、 $1 \sim n$  までの番号が割り振られている場合を考え、番号が小さいものほど優先順位は高いものとする。また、プレファレンス値が取りうる値を  $n$  個の区間に分割し、プレファレンスの昇順に  $1 \sim n$  の番号を割り振る。さらに、区間  $i$  のプレファレンス値を持ったデータグラムを  $i$  のキューに割り振る。以上の処理をバッファ書き込み制御部において行うことにより、小さいプレファレンス値を持つデータグラムほど優先度の高いバッファメモリに割り振られることとなり、この発明の効果を実現することが可能となる。

## 【 0 0 4 4 】

この発明では、通常のデータグラム（例えば IP）のヘッダに存在するデータグラムのサイズフィールドをプレファレンス値として利用することができる。これに対し、以下に示す例では、トラヒック観測装置 8 においてトラヒックの観測により計算された値をデータグラムのヘッダの特別なフィールドに挿入することを基本としている。しかし、この例ではデータグラムのフォーマットに特別なフィールドは必要ない、但しデータグラムのサイズは網がつけるものではなくユーザ端末がつける値であるため、トラヒック観測装置 8 は、書かれているデータグラムのサイズの値と実際のサイズが正しいかどうかというテストを行い、正しくない場合にはそのデータグラムを廃棄する処理を行う。

## 【 0 0 4 5 】

また、プレファレンス値の他の計算機能として、一つ前のデータグラムの送出時刻と、現在時刻との差の逆数（データグラム転送間隔の逆数）を用いることができる。この場合には、現在時刻を  $t$ 、現在到着したデータグラムの大きさを  $M$  とし、また、一つ前のデータグラムの送出時刻を  $t_i$ 、データグラムの大きさを  $M_i$ 、 $\alpha$  を影響度の度合を調整する定数とすると、 $\alpha (M/M_i) \{ 1 / (t - t_i) \}$  式で表現される値を用いる。この値は、瞬間的なユーザのトラヒックの速度を意味する値であるため、この値が大きいほど網に対するインパクトが大き

いと判断することが可能である。

【0046】

さらに、プレファレンス値の他の計算機能として、データグラムのサイズおよび連続するデータグラムの間隔から計算する平均レート（スライディングウィンドウ方式）を用いることができる。この場合には、現在から  $i$  個以前のデータグラムの大きさを  $M_i$ 、到着時刻を  $t_i$ 、遡るデータグラム数を  $n$ 、影響の度合を調整する定数を  $\alpha_i$  とすると（ $n$ 、 $\alpha_i$  はあらかじめ設定された値）、平均レート  $V_K$  は、 $\alpha_i M_i / (T_i - T_{i+1})$  の  $i$  が  $0 \sim n$  までの和として求められる。

【0047】

この値  $V_K$  は前記データグラムの転送間隔の逆数を用いた場合より長い時間間隔でみた平均レートとなるので、やはり網に対するインパクトを評価することが可能である。

【0048】

また、プレファレンス値の他の計算機能として、データグラムのサイズおよび、連続するデータグラムの間隔から計算する観測期間の平均レートを用いることができる。すなわち、観測期間を  $T_A$  とすると、この観測期間  $T_A$  に到着したデータグラムのサイズの和を、観測期間  $T_A$  で除算したものを利用する。また、データグラムのサイズの和そのものを用いることも可能であり、また、その和をアクセス網の物理リンク速度より決まる和の最大値で除算し、正規化を行った値とすることも可能である。

【0049】

また、プレファレンス値の他の計算機能として、ユーザが送出したデータグラムの個数と、受信したデータグラムの個数の差を用いることができる。通常、データグラム転送網を適正に用いる場合には、データグラムの配達が保証されないため、ユーザ端末は正しく受信できたデータグラムに対し、正しく受信できたことを送信ユーザ端末に知らせる ACK データパケットを送信する。従って、このような適正な方法を用いてデータグラム通信網を利用する場合には、送信したデータグラム数と受信したデータグラム数はほぼ等しくなる。一方、この発明が課

題とするような方法をユーザがとる場合には、網内で多くのデータグラムが廃棄され、それに対するACKデータパケットが送信されないため、送信データグラム数と受信データグラム数には大きな開きができると考えられる。従って、この差をプレファレンス値とすれば、適正な利用方法を行っているものがより優先的に転送されることとなり、この発明の効果を得ることが可能となる。

【0050】

【発明の効果】

以上のように、この発明によれば、データグラム転送／廃棄方法を実現するデータグラム転送装置を用いることで、ユーザが多くの網資源を獲得しようとして必要以上のデータグラムを送出することを抑制でき、これによって輻輳崩壊状態に陥らない安定したデータグラム通信網を実現できるという効果が得られる。

【図面の簡単な説明】

【図1】 この発明の実施の一形態によるデータグラム転送システムを示すブロック図である。

【図2】 図1におけるトラヒック観測装置の詳細を示すブロック図である。

【図3】 この発明におけるデータグラムのヘッダフィールドを示す説明図である。

【図4】 この発明における出力インタフェース部内のバッファ書き込み制御部を示すブロック図である。

【図5】 この発明における書き込み制御機能によるバッファメモリへの書き込み手順を示すフローチャートである。

【図6】 この発明における書き込み制御機能によるバッファメモリへの他の書き込み手順を示すフローチャートである。

【図7】 この発明における書き込み制御機能によるバッファメモリへの他の書き込み手順を示すフローチャートである。

【図8】 この発明における書き込み制御機能によるバッファメモリへの他の書き込み手順を示すフローチャートである。

【図9】 この発明における書き込み制御機能によるバッファメモリへの他

の書き込み手順を示すフローチャートである。

【図 1 0】 この発明における出力インタフェース部内のバッファ書き込み制御部の他の例を示すブロック図である。

【図 1 1】 この発明における出力インタフェース部の他の例を示すブロック図である。

【図 1 2】 従来およびこの発明のデータグラム転送装置の基本構成を示すブロック図である。

【図 1 3】 図 1 2における入力インタフェース部の従来例を示すブロック図である。

【図 1 4】 図 1 2における出力インタフェース部の従来例を示すブロック図である。

【図 1 5】 従来のファーストインファーストアウト方式によるデータグラム転送手順を示すフローチャートである。

【図 1 6】 従来のランダムアーリーディテクション方式によるデータグラム転送手順を示すフローチャートである。

【符号の説明】

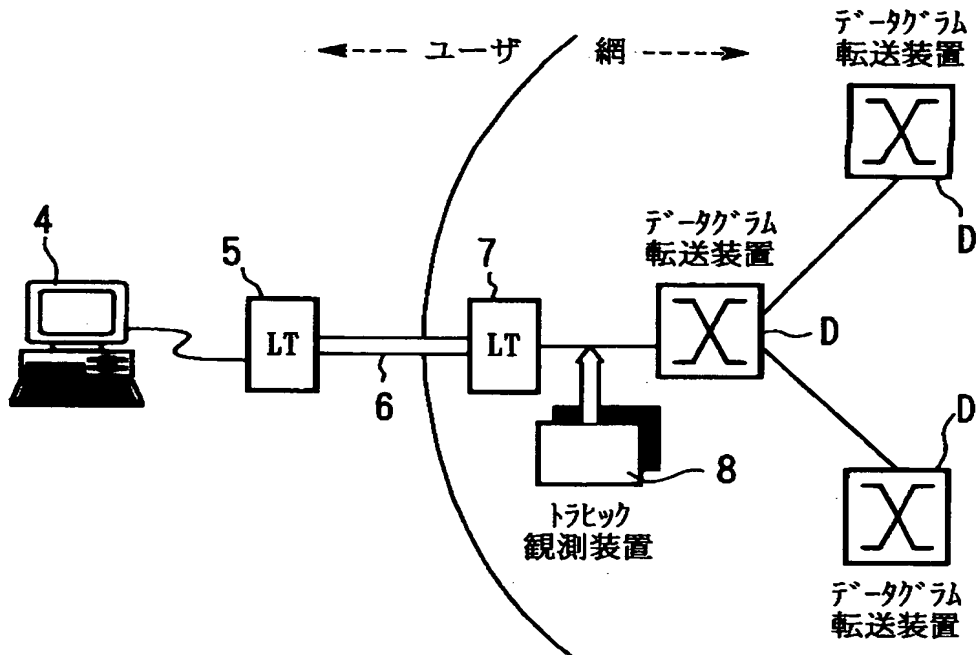
- 1 入力 I / F 部 (入力インタフェース部)
- 2 出力 I / F 部 (出力インタフェース部)
- 2 c バッファ書き込み制御部
- 2 c 4 プレファレンス値保存機能
- 3 バックプレーンスイッチ部
- 8 トラヒック観測装置
- 8 a トラヒック観測機能
- 8 b プレファレンス値計算機能
- 8 c プレファレンス値挿入機能



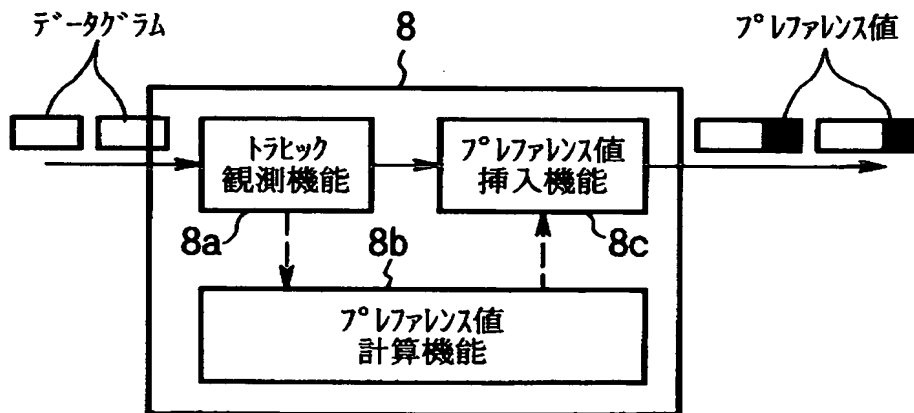
【書類名】

図面

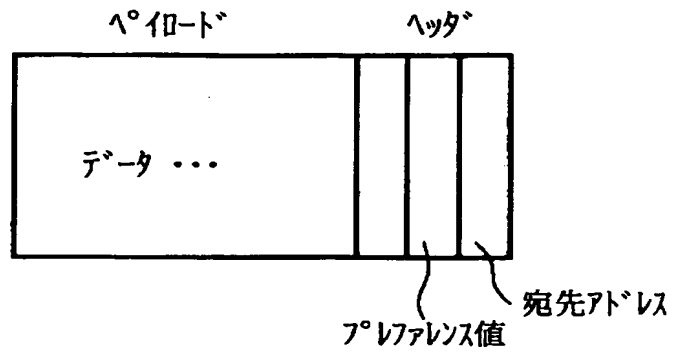
【図 1】



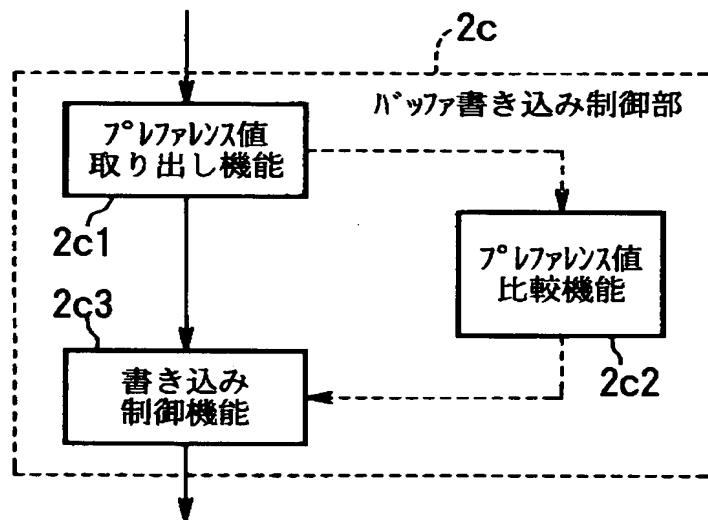
【図 2】



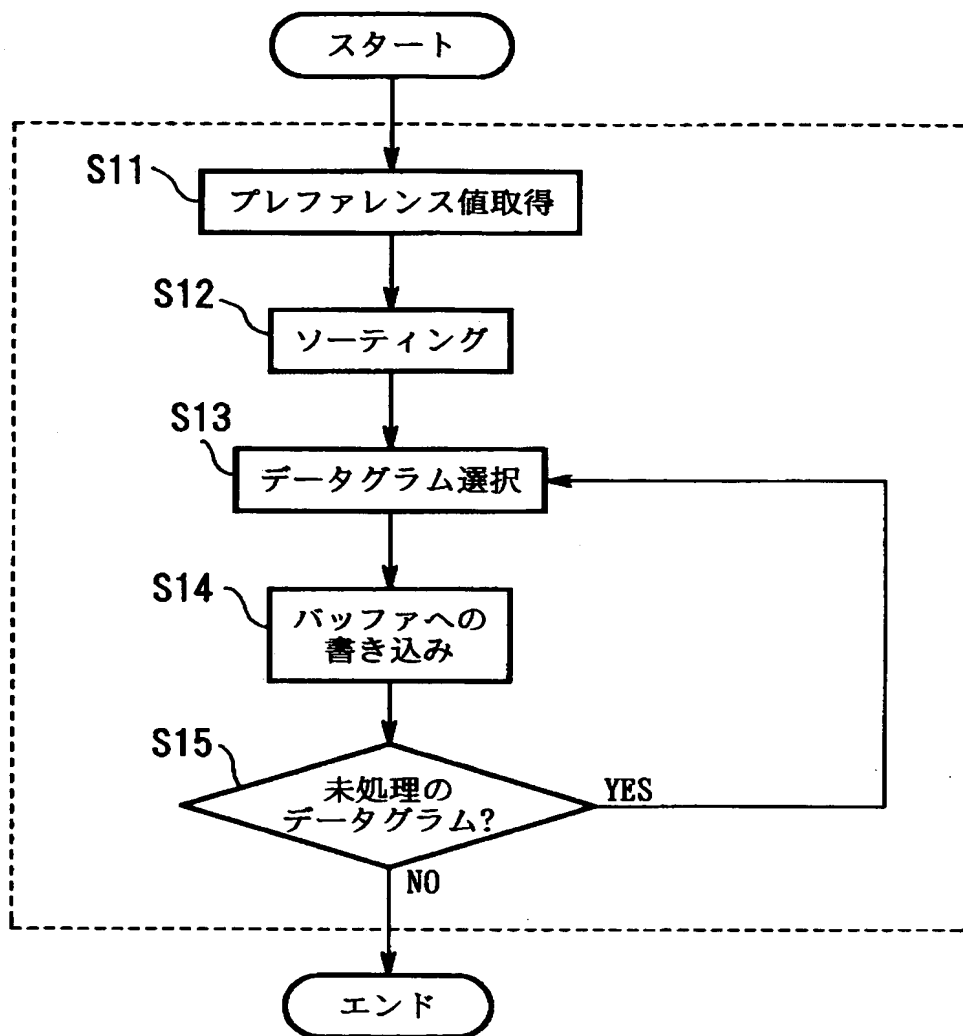
【図 3】



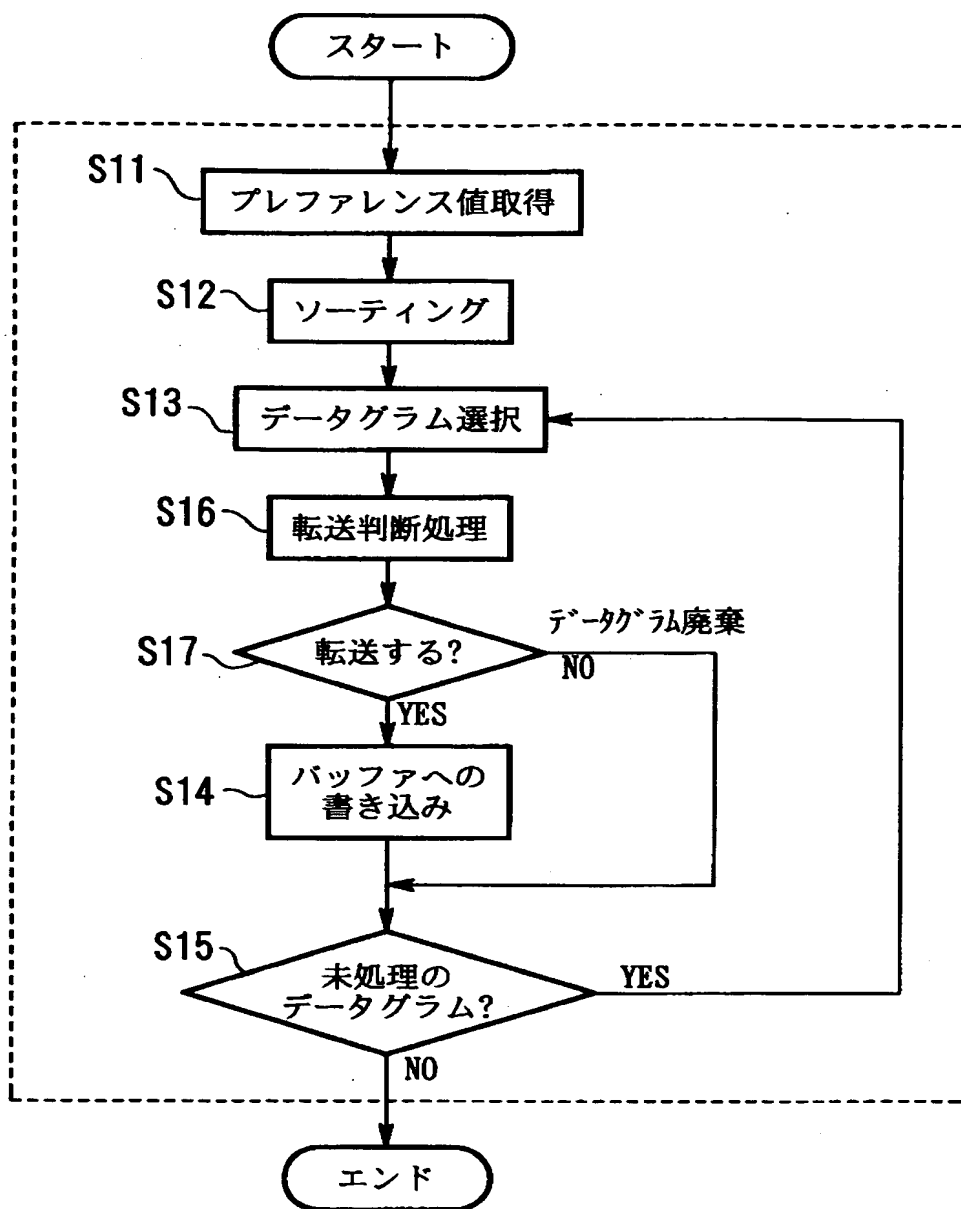
【図 4】



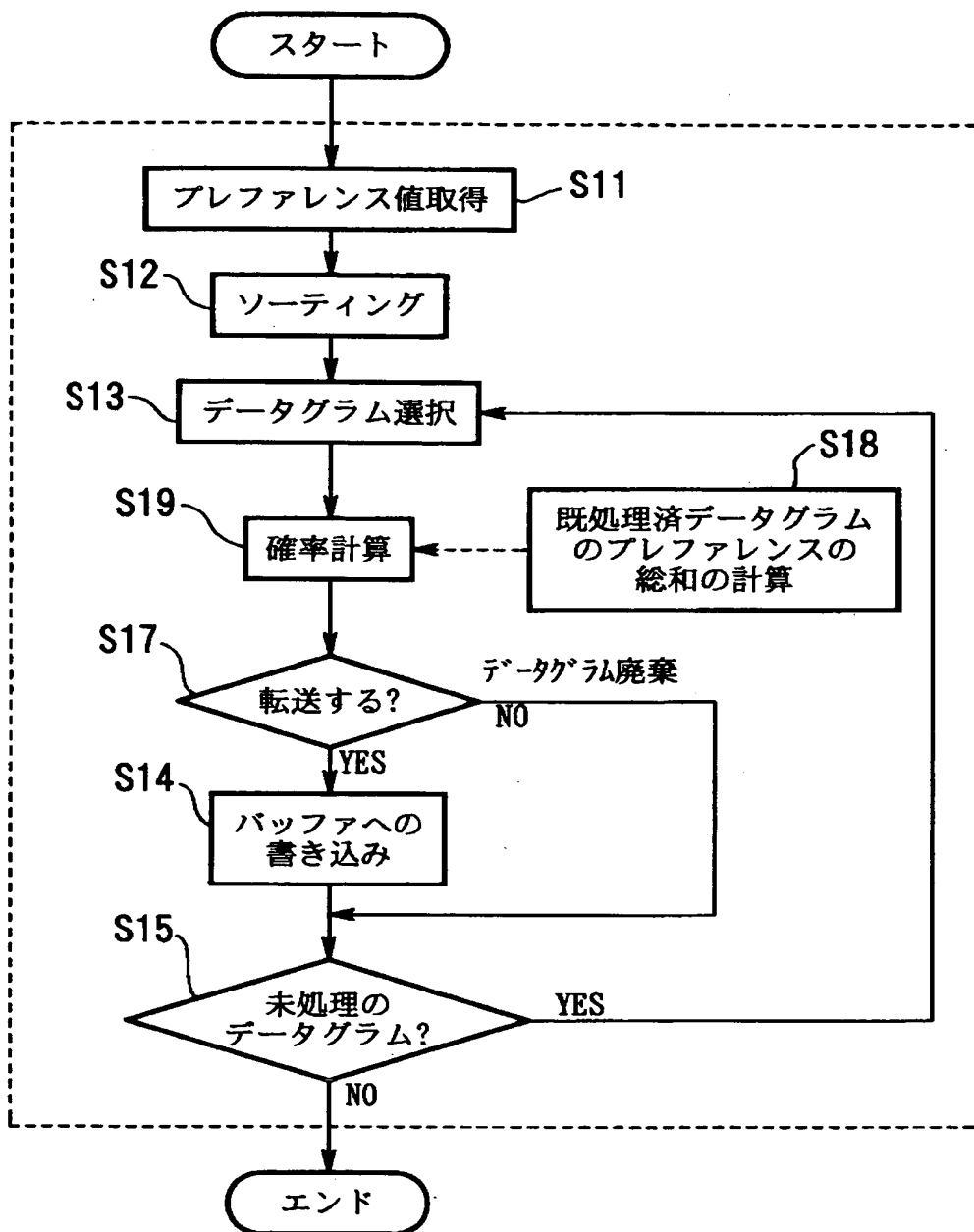
【図 5】



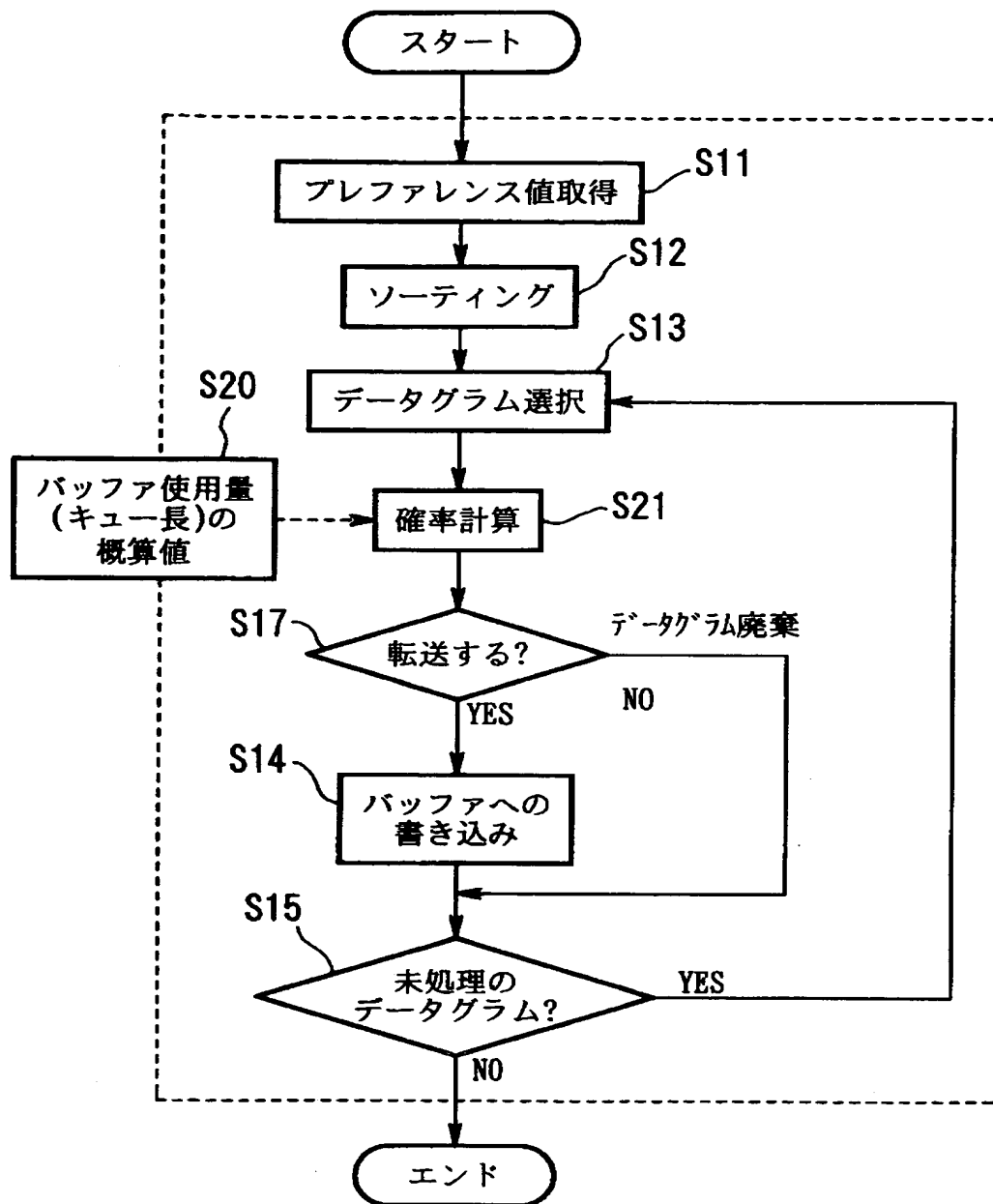
【図6】



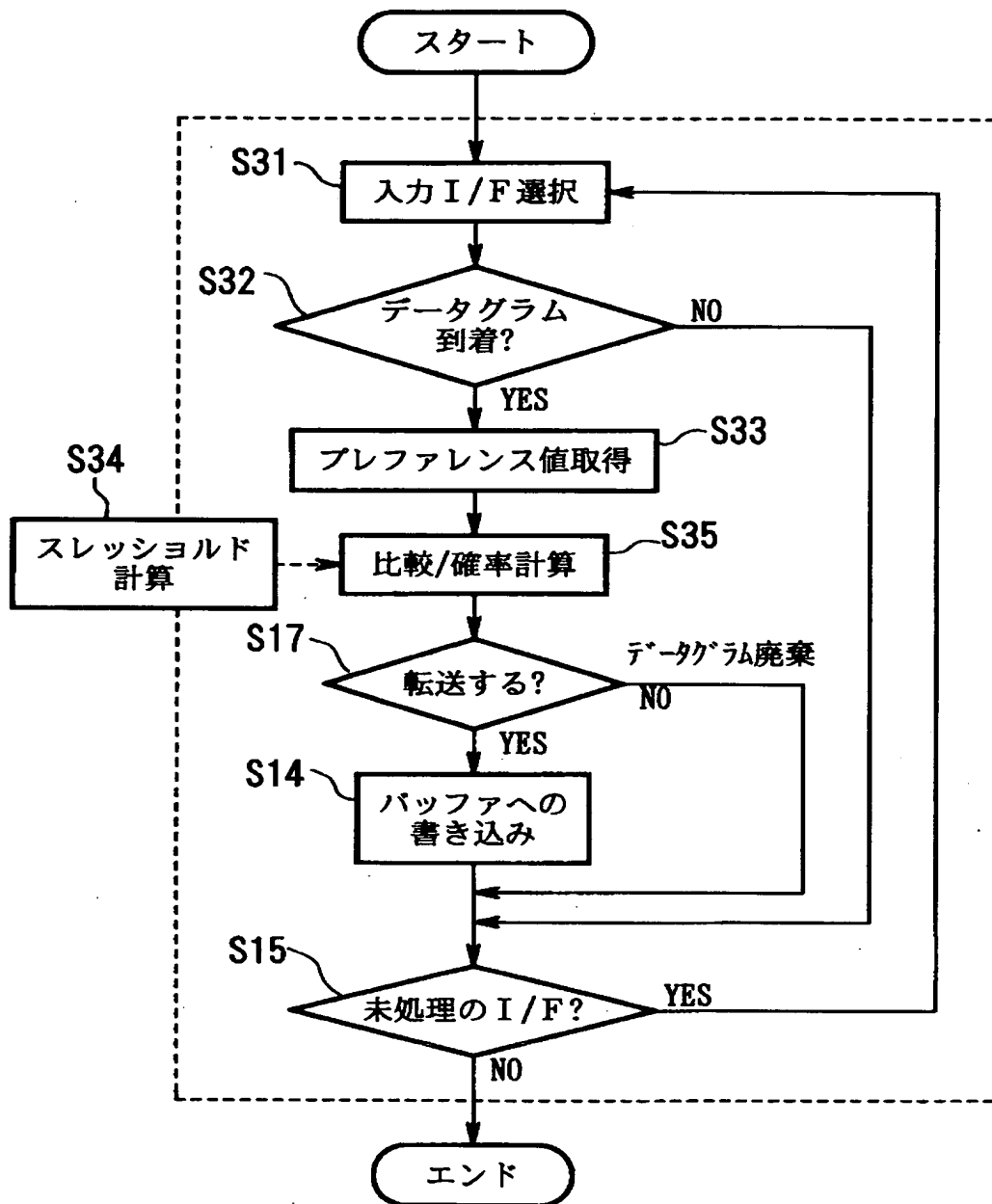
【図 7】



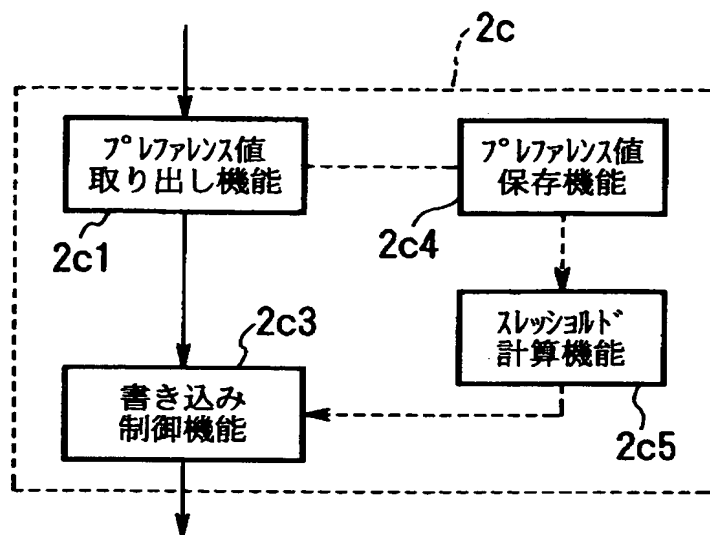
【図 8】



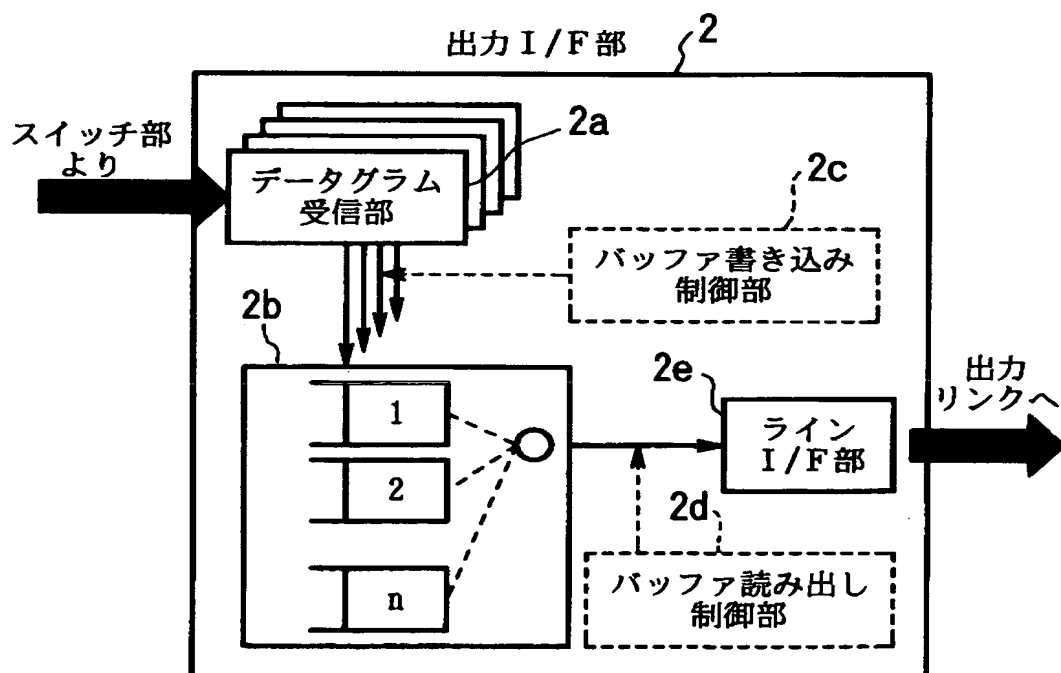
【図9】



【図 1 0】

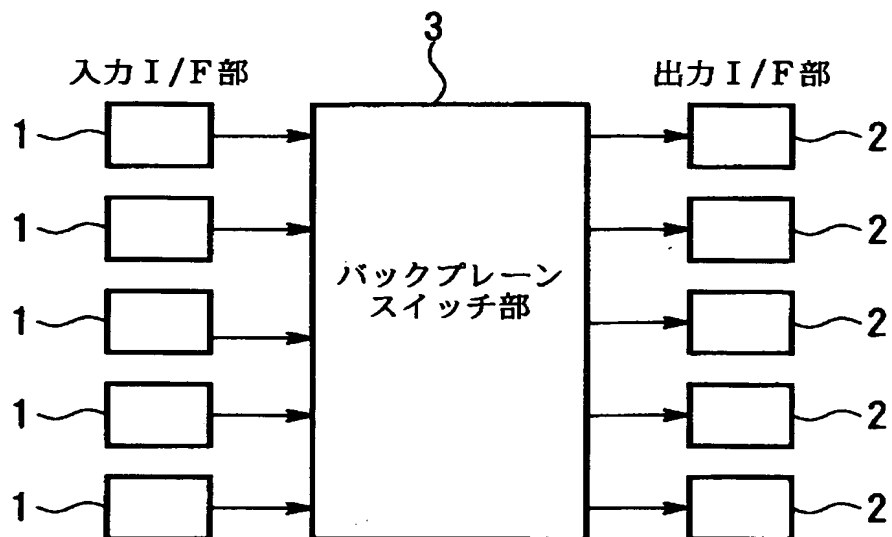


【図 1 1】

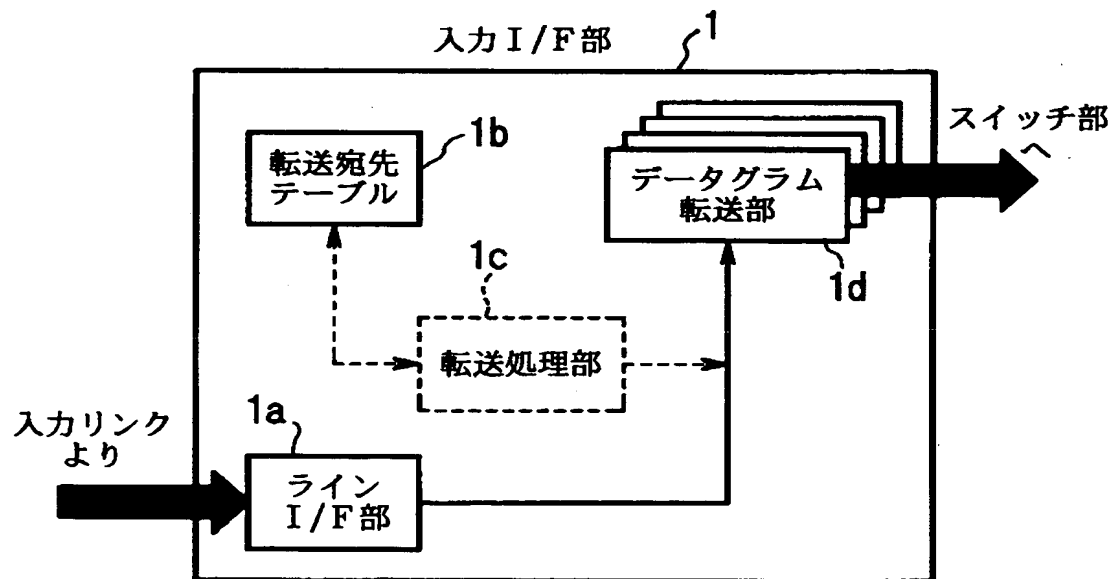




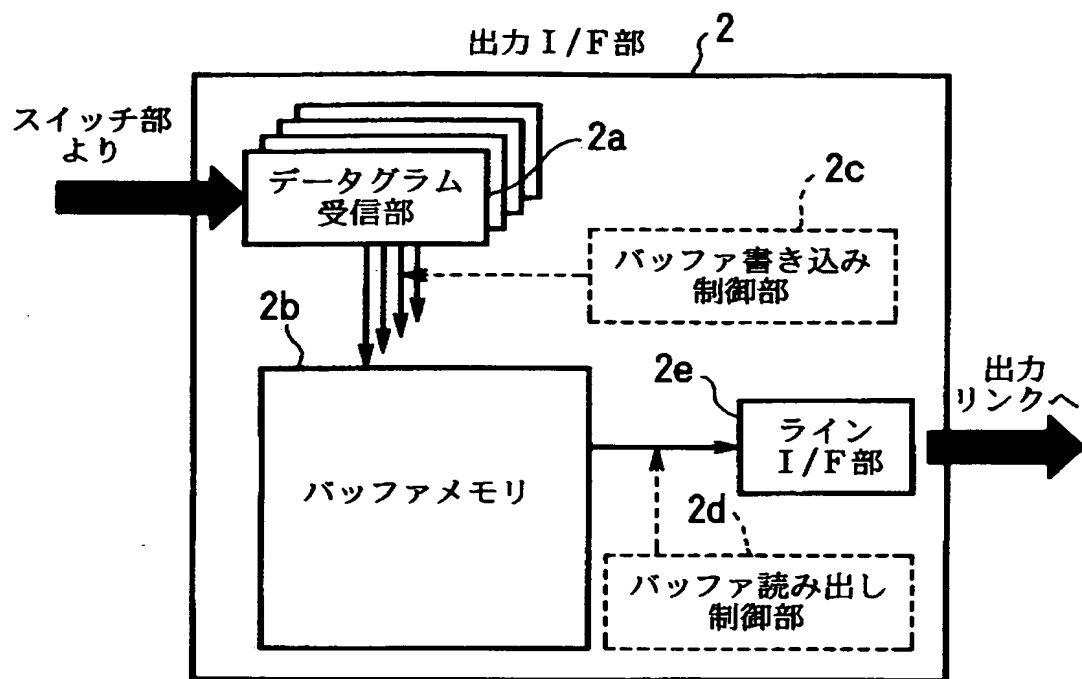
【図 1 2】



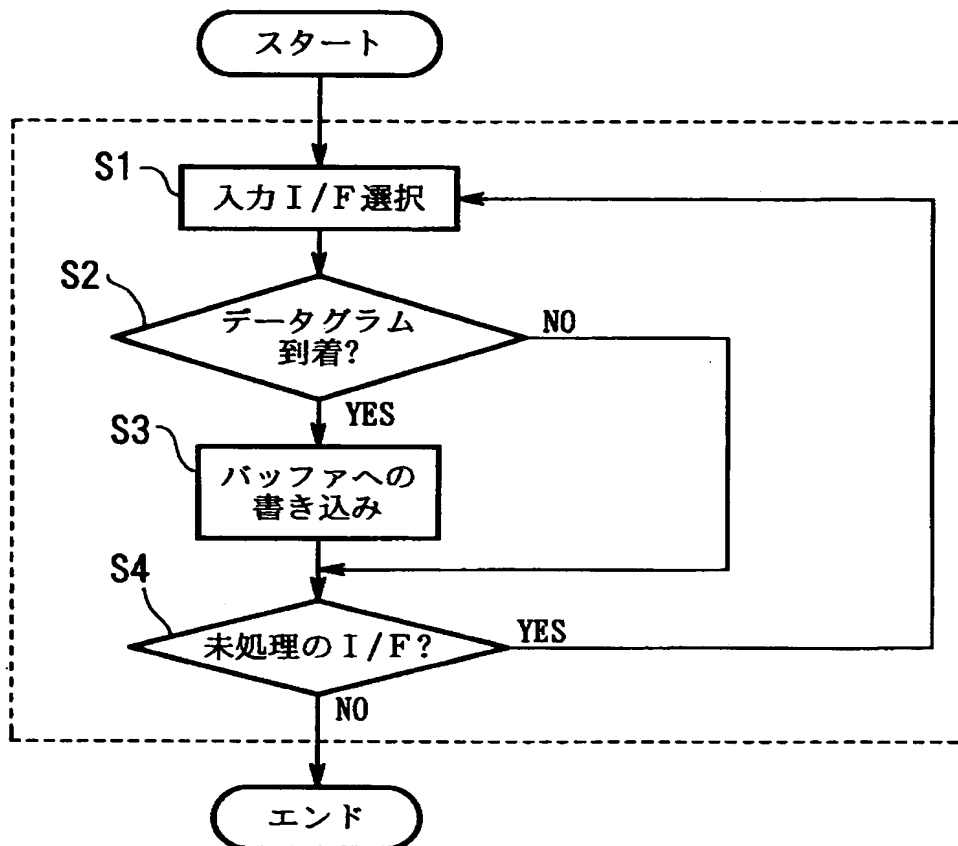
【図 1 3】



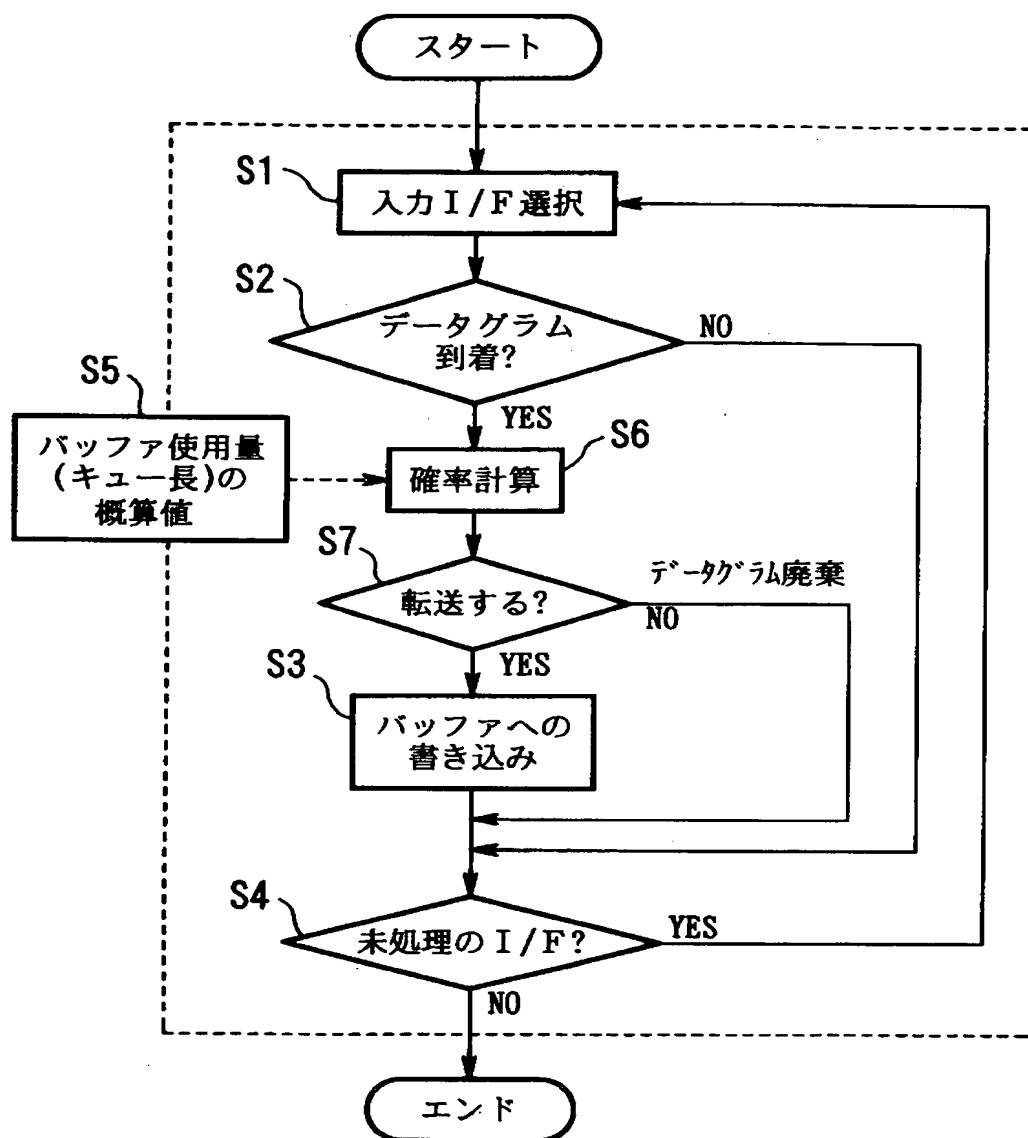
【図 14】



【図 1 5】



【図 16】



【書類名】            要約書

【要約】

【課題】 データグラムの送出に対する網の評価により得られる情報を利用して転送処理を行うことで、輻輳崩壊状態に陥らない安定したデータグラム通信網を実現可能にする。

【解決手段】 データグラムのヘッダに記載された転送宛先アドレスに向ってデータグラムをリレーするデータグラム転送システムにあって、データグラムに関するトラヒック情報に基づいてユーザのトラヒックの網に対するインパクトを評価し、その評価値をヘッダに書き込んで、この評価に従った順序による転送処理を行わせるトラヒック観測手段 8 を設ける。

【選択図】            図 1

出 願 人 履 歴 情 報

識別番号 [000004226]

1. 変更年月日 1995年 9月21日  
[変更理由] 住所変更  
住 所 東京都新宿区西新宿三丁目19番2号  
氏 名 日本電信電話株式会社
2. 変更年月日 1999年 7月15日  
[変更理由] 住所変更  
住 所 東京都千代田区大手町二丁目3番1号  
氏 名 日本電信電話株式会社